DOCUMENT RESUME

ED 430 278 EA 029 794

AUTHOR DeBray, Elizabeth H.
TITLE Incentives in States' Educational Accountability Systems: Is the Assumption of Continuous Improvement Included?
PUB DATE 1999-04-00
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Accountability; Change Strategies; Educational Change; *Educational Improvement; Elementary Secondary Education; *Excellence in Education; School Effectiveness; *State Action

ABSTRACT

This paper discusses the designs and effects of emerging state-level accountability systems. It claims that in a few years' time, state-level policymakers will be faced with recognizing that a substantial number of schools in their states will have failed to progress toward academic standards at the rate their reform plans demand. The paper has three sections. The first part examines some of the theories around state-level policies and performance-based accountability systems and places emphasis on the role of rewards and sanctions in improving instruction. The second section, by drawing on examples of accountability systems in three states (Kentucky, Maryland, and Mississippi), illustrates how the problem with schools in the "middle" (those that are neither failing nor producing high outcomes) is seen to emerge under different accountability policy regimes. The final section offers some possible future directions, such as differentiating and targeting policies, for thinking about incentives for continuous improvement under accountability policies. The article considers some common assumptions about school governance and education policy that have emerged in the 1980s and 1990s and that underlie accountability systems and uses these assumptions to ask questions about their effects on schools in the middle. Contains 35 references. (RJM)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

EA

ED 430 278

# Incentives in States' Educational Accountability Systems: Is the Assumption of Continuous Improvement Included?

**Elizabeth H. DeBray**
**Harvard Graduate School of Education**

Paper presented at the
Annual Meeting of the
American Educational Research Association,
April 19-23, 1999
Montreal, Canada

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

E. DeBray

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)



EA 029 794

BEST COPY AVAILABLE

*Introduction*

My purpose in this paper is to frame an argument about the design and effects of emerging state-level accountability systems.1 With increasing frequency, states are setting in place policies whose purpose is to monitor schools' progress toward goals, and allocate rewards and sanctions to them according to their performance.

I posit that in a few years' time, state-level policymakers will be faced with a dilemma. A substantial body of schools in their states will have failed to progress toward academic standards at the rate the reform plans demanded. These schools likely would not be those plagued by serious academic failures, nor be candidates for reconstitution or other state intervention. In fact, the communities in which they are located might not question their academic performance; achievement test scores might cluster around the state median. These "schools in the middle" in fact might have reason to have been celebrated by policymakers at one time in the early years of the accountability plan; perhaps they served students of low socioeconomic status and they made substantial gains. Or these schools might serve suburban communities whose students and teachers have produced solid test scores and do not experience the pressure to improve. *These schools have not shown the continuous growth that their state's policy envisioned.* Depending on a particular state's policy design, we would view this non-progressing middle range differently -- relative to other schools grouped within the same socioeconomic band, or relative to the rate at which other schools are progressing (discussed in greater detail in section II). Whatever their progress has been in the past, or what their demographic characteristics are, the commonality across these schools is that after a certain time, they will likely turn up neither on the state's list for a cash reward, nor as a candidate for intervention or sanction.

Yet policymakers ultimately will need to know: if the level of teaching and learning in these schools reflects mediocrity, how can this broad middle be reached? I argue that a precursor to conceptualizing policy design is thinking about these schools as a group with particular needs and problems. Only then could states target incentives to encourage them to progress beyond their stasis.

An assumption underlying many of these state-level systems is that all schools can continuously improve their performance. There are two potential problems, however, in the design of these policies. The first is that the design addressed the middle range, but the incentives are weak to spur school-level improvements. The second is that the policy design is, in fact, "under-specified," and never addressed the problem of continuous improvement for schools in the middle. I will argue that the incentives built into most policies do not address continuous improvement, particularly for schools that are in the broad "middle": that is, they are not failing, nor are they producing high outcomes. This incentive problem has rarely been written about in the accountability literature, though some authors have referred to it in passing as a possible flaw in design. Elmore, Abelmann, and Fuhrman (1996, p.80), discussing state-level accountability systems in

---

1 I wish to thank Susan Fuhrman and Charles Abelmann for encouraging me to pursue the topic; and Richard Elmore for his feedback during his seminar on Issues in Large-Scale Instructional Improvement, and on my various drafts.

Mississippi and Kentucky, note: "In both states, certain elements of the systems raise questions about the strength and effectiveness of the incentive structure. One important issue is the extent to which the systems focus only on the top and bottom of the performance distribution, leaving out those schools and districts falling in between."

It is quite likely that in the near future, policymakers will seek to "fine-tune" accountability systems -- that is, rethink how many current school improvement policies operate, and assess the extent to which they have reached most schools. If research can shed light on the broad middle range of schools, and how they fare under these systems, future policies might be designed with their particular needs in mind. Recent research by Newman and Rigdon (1997), and by the Consortium for Policy Research in Education (Elmore, Abelmann, Marshall, and Even, 1999) begins to highlight the variation of policy responses, dependent upon schools' internal accountability systems and how they interface with external policy regimes.

This paper has three sections. In the first part, I will examine some of the theories around state-level policies and performance-based accountability systems, and from these writings, extract my own questions about possible implications for schools in this "middle" range. (Three of these are deregulation and continuous improvement; classification of schools; and what is known about the effects of rewards and sanctions.) In the second section, drawing on examples of the accountability systems in three states (Kentucky, Maryland, and Mississippi), I will illustrate how the two potential problems around schools in the middle described above are seen to emerge under different accountability policy regimes. In the final section, I offer some possible future directions for thinking about incentives for continuous improvement under accountability policies.

Accountability policies of greater sophistication could be directed toward building instructional capacity, and moving away from thinking primarily about rewards and sanctions. Fuhrman and Elmore write about the implications of deregulation in state policies, captured the importance policymakers' starting to think about differentiated responses (1992, pp. 27-28):

> This means less reliance on mandates and incentives and the bundling of a variety of instruments to achieve particular goals. If a goal is important, the range of local response might be anticipated through the use of a variety of instruments that speak to the distribution of local needs, priorities, and capacities...developing complex, differentiated policy approaches itself is a highly sophisticated endeavor which might require more capacity than many state legislatures or agencies possess.

Next, I consider some common assumptions about school governance and education policy that have emerged in the 1980s and 1990s that underlie accountability systems, and use them to ask questions about their effects on schools that are neither failing nor achieving at high levels (or, under particular policies, it could be said that the schools are neither declining nor improving).

## I. Looking at the Literature on Performance-Based Systems and State Policies to Derive Questions about the Middle

*Deregulation, Continuous Improvement*

Two policy ideas that are embedded in many state accountability systems are those of *continuous improvement* and *flexibility* (i.e., less state intervention in exchange for improved school outcomes). I include them because in states where a middle range has not been addressed by the accountability system, these two ideas may be the invisible policy provisions that are widely assumed will help all schools progress. Yet research has not established a credible link between state deregulation and better teaching and learning; nor whether schools know how to utilize information and results to continuously improve their performance. I am suggesting that in many schools with average achievement, neither of these policy ideas have become targeted enough to stimulate specific improvement strategies.

Two national-level commissions that endorsed states' regulating schools less if they could produce better results were the National Governors' Association's *Time for Results* in 1991; and *Putting Learning First* by the Committee for Economic Development (1994). Both reports emphasized that an appropriate role for states is to define academic outcomes and levels and performance goals for students, but then allow school leaders the flexibility to meet these goals. For instance, the Task Force on Teaching in *Time for Results* recommended:

> State and local authorities can deliberate with the educators and then be explicit about expected levels of academic performance. Then they should allow teachers, administrators, and parents to devise ways to meet these levels. Solutions are not obvious here. It is not a matter of defining the courses students must take, but a painstaking and continuing inquiry into what skills students should have... (38).

The governors' proposition was that if states and districts set clear academic goals, then schools and teachers should be given the freedom to meet the goals any way that they want. Their "action agenda" recommended "reduced state requirements that limit the ways in which local districts and individual schools help their students achieve the expected levels" (National Governors Association, 1991, p.95).

Three years later, the Committee for Economic Development (1994) endorsed school governance that matched goals for achievement with flexible, decentralized management:

> We believe that compliance and control must be replaced by more flexible management that gives more authority and accountability for results to teachers, administrators, parents and students. This "flatter" management structure must be coupled with a variety of incentives, focused on measurable academic achievement, that will motivate improved performance (Committee for Economic Development, 1994, p. 11).

I suggest that the supposed boon of deregulation that has been invoked by these commissions might have been expected to assist middle-performing schools in particular. High performing schools were likely producing good outcomes absent state involvement, while failing schools may have already been monitored by the state (in some cases).

Yet it is important to notice that deregulation and flexibility, while embedded in most states' accountability systems, are primarily ideas about management, not about instruction. Whether this diffuse policy direction has produced or may produce the conditions for improved instruction is uncertain. Brian Rowan, in "Standards as Incentives for Educational Reform" (1996), argues that there are still many reasons why the effects of standards on teacher performance have been weak. One reason is that "teachers in the United States have great freedom in choosing learning goals for students, even in school systems that have developed elaborate curricular guides and grade-level expectations for student learning" (1996, p. 205). Another reason that greater school autonomy, even when coupled with fairly specific state standards, may not produce the desired changes in instruction, is that the outcomes are not, in Rowan's words, "meaningful to teachers":

> ...elaborate and formalized student standards often fail to be meaningful to teachers for two reasons. First, the outcomes described in such standards often are not those that teachers personally value. In addition, school systems rarely reward or punish teachers based on achievement of these standards (Floden et. al., 1988). As a result, many of the elaborate goal-setting strategies used in American education have only a modest effect on instructional practices (Rowan, 1996, p. 205).

When looking at states' systems, we could ask with respect to the middle: is setting district or state-level goals and then lifting some requirements enough to spur schools to raise their own levels of expectations for students? Based on the literature, it seems that educational policy research has not answered this question.

*The Theory of Continuous Improvement*

"Continuous improvement" is an idea originating in the literature of management and organizational theory, not education, but many in the educational sector have embraced the concept. It is interesting that while deregulation focuses on outcomes, continuous improvement outlines a *process* of organizational knowledge utilization:

> High-performing organizations are able to accomplish their mission by continually improving their capacity to deliver highly valued outcomes to their stakeholders, and in return to continue to receive the resources required for ongoing performance. This view of high performance combines some of the classic elements of the definition of an effective system. Organizations have to accomplish their mission and they have to provide products and services that are valued by the stakeholders in their environment so that they have access to an ongoing stream of the resources necessary for their own survival (Katz and Kahn, 1978, Pfeffer and Salancik, 1978) (in Mohrman and Wohlstetter, 1994, p. 5).

Perhaps the most widespread policies growing out of the "common wisdom" of continuous improvement has translated into delegation of decision-making power to school-site councils. Summers and Johnson (1996) outline several arguments for decentralization in K-12 education, including negative externalities and inefficiencies in governmental regulation of a heterogeneous student population; and citizens being able to better express their individual preferences and improving the chance that the organization can respond. However, the literature on the effectiveness of site-based management has been inconclusive about its effects on students' learning and achievement (1996, p. 80). Summers and Johnson note that studies of site-based management have not linked decentralization to learning outcomes:

> There is an implicit assumption that, if the processes of decision making change, schools will be more effective instruments for educating children. The studies were designed, however, to look at the effects of SBM on governance processes, not educational outcomes, just as SBM efforts are designed to alter stakeholder relationships via governance changes, not to change student performance.
>
> Essentially, the large literature on the effectiveness of SBM ignored the effects on student achievement, either because the SBM advocates do not regard achievement as an important output measure or because there is faith that increased school discretion will increase student learning. As a result, there is little evidence to support the notion that SBM is effective in increasing student performance. There are very few quantitative studies, the studies are not statistically rigorous, and the evidence of positive results is either weak or nonexistent (Summers and Johnson, 1996, p. 80).

When applied to schools, the theory is that teachers and administrators will use information about performance and achievement to improve. For instance, in Maryland, the state education agency publishes a "red book" annually, containing detailed information on a variety of indicators about schools in each county, including assessment results. According to state department leaders, the state assumes that the information will be utilized by county and school-level administrators in targeting areas for improvement (interview with Richard Steinke, 11/4/97). After almost six years of this publication of scores on statewide assessments, however, there is no school in Maryland which has achieved the performance level of "satisfactory," as the state board has defined that goal. It raises the question of whether merely making information widely available to the public is enough to produce the desired incentive of galvanizing school leaders to improve teaching and learning. It also highlights the issue of whether the state's rewards and sanctions program is reaching schools which are neither declining nor failing, but are not attaining the "satisfactory" level. The theory may be that schools will utilize knowledge, but whether they have the capacity to do so, or even experience it as a strong incentive in and of itself, is unknown.

*Value-Added: A Policy Approach that Accounts for the Middle*

Policies that are based on the concept of "value added" are the ones most explicitly accounting for middle-performing schools and their continuous improvement.

State accountability programs in South Carolina and North Carolina are attempting to evaluate how schools serve students in relation to a measure of their predicted performance (Ladd, Roselius, and Walsh, 1997). Since each school is supposed to be rewarded commensurate with its predicted level of achievement (calculated based on a number of different indicators about students, community, and family background), schools with achievement around the state median level would not be overlooked. Unlike some other systems, which reward most improved schools or sanction low-performing or declining schools, a value-added plan would compare each school, using regression analysis, to its own statistically predicted level of performance. The state must then make a decision about its terms for continuous improvement: in what time frame must schools meet certain targets for improving their contribution to students' education?

What makes this policy a bit different from others is that when implemented fully, there would be no middle-performing group of schools. The state would maintain indicators about the inputs and resources of individual schools, and determine whether or not a school had met, exceeded, or fallen short of its predicted level of performance. It is at the state's discretion what level of performance to recognize for rewards and sanctions. In North Carolina, an effective school is one in which the school is between target level (its predicted performance level) and at least 10% above its predicted performance level (Ladd et al. 1997). Since the state is not ranking schools based on absolute performance, but rather is evaluating whether schools have achieved specific predicted target goals, the distinctions of top, middle, and bottom ranges are not as readily apparent.

However, states could devise a variety of policy designs to accompany the implementation of a value-added model of school effectiveness. Once they are able to measure where schools are relative to their predicted performance, they would want to tie specific incentives to meeting those goals. For instance, for how many years would a school have to fall short of its predicted performance level without sanction? What about schools that meet target performance goals during the first several years of a reform, but then their gains level off? The point is that even under this model, states may have will still ultimately face questions about incentives for continuous improvement.

There are several initial questions to ask about the design and use of information about the use of value-added measures of effectiveness in state accountability policy. Which indicators are included in the regression equation determining schools' predicted performance? How can information from the value-added program could be communicated back to schools and used for instructional improvement? That is, is there any state intervention based on the findings about where schools stand compared to where they might be?

**Rewards and Sanctions: How Effective Ultimately as Instruments For Improving Instruction?**

State accountability policies generally rely on the allocation of rewards and sanctions for schools' (or in Mississippi's case, districts') performance or improvement. What is known about the effectiveness of these reward programs? Based on states' experience so far, do we know how responsive schools and districts are to rewards and

sanctions? Are there limits to schools' improvement even when rewards are offered; and does focusing on top performers and failing schools create bifurcation, overlooking a broader group of schools which do not feel the pressure of incentives?

Rewards and sanctions are by definition essential components of an accountability system -- unless there are consequences for schools' improvement, and state responses to student outcomes, then there is no such system. A question I pose for further research is: how are these incentives experienced in middle-performing schools and districts? Do school officials know what kind of performance or improvement is needed to qualify for a reward? I am suggesting that policymakers should recognize that financial rewards are but one type of policy strategy, and that offering them may only yield "short-term rewards" (McDonnell and Elmore, 1990).

Also, policy-makers should consider the effects of a prominent focus on reconstitution and low-performing schools on *other* classes of schools within the same state. In Maryland, for example, the School Performance Index (SPI) and Change Index (CI) are generated primarily to determine "reconstitution eligibility" (interview with Richard Steinke, 11/4/97). For those schools that have never been reconstitution-eligible, how is this program experienced? I would argue that a policy oriented toward the lowest tier of failing schools is not generating incentives for continuous improvement in other schools.

How schools are classified will often determine not only access to rewards and sanctions, but also access to other capacity-building resources. I turn next to this problem.

*Classification: An Often-Overlooked Design Issue that Will Determine Treatment*

The critical design feature in state policies that determines how rewards, sanctions, and incentives will be focused in an accountability system is classification of schools or districts. For example, David Cohen notes the problem of how states set thresholds for top and bottom performing schools -- and in doing so, notes that this is a matter that has been discussed little in policy design. Should any school that does not receive a reward be labeled "failed"? This would produce certain problems, he writes (1996, p.81):

> ...it would label schools as either outstanding or awful rather than also being satisfactory or indifferent. Additionally, if the criteria for success were set relatively high, a pass-fail approach could produce a politically unacceptable avalanche of failures, with the likely result that criteria for success would be abandoned or set lower.

Yet then it is unclear what policymakers should do about designating middle-performing or satisfactory schools:

> Perhaps only especially low-performing schools should be rated as unsatisfactory, with those in the middle left unclassified. Such decisions would be consequential, for designating schools as `failed' or `unsatisfactory' might do more damage than the label of success could do good. Decisions about where to draw the line for failure thus would raise all the issues concerning criteria of success that I just

> discussed, perhaps with even greater stakes. There has been little discussion of this matter (Cohen, 1996, p. 81).

The issue of designation of performance is also important because under some state accountability plans, schools designated as "in decline" or "failed" may be eligible for more resources. Schools' classification may ultimately determine some of their access to additional resources or technical assistance. Legislation passed in Florida, for instance, requires districts and schools for provision of extra academic help to students who are not proficient in reading, writing or math in grades 1 through 5 and instructional assistance for students who do not pass any section of the high school exit examination. The legislature has appropriated funds for this extra instructional assistance (American Federation of Teachers, 1997). In Kentucky, schools that have not made the required gains every two years are assigned distinguished educators to assist the school with its instructional programs, and increasingly, staff are claiming that their presence is a bonus, and their presence is not seen as a sanction (Kelley, 1997). Policy researchers have also found that states have targeted resources on low-performing schools and school districts. For instance, in Michigan, annual grants of up to $60,000 were provided to eleven low-achieving, urban, or extreme rural school districts through the Statewide Systemic Initiative (Goertz, Floden and O'Day, 1995). Another example is that a high percentage of a state's Obey-Porter (or comprehensive school research demonstration) monies must be given to high-poverty schools.2

Such capacity-building interventions in low-performing schools are an important development in state education policy. *These interventions raise an important research question about what kinds of capacity-building are similarly available to middle-performing schools. As the designations of "low-performing" or "failed" proliferate, it will be necessary to think about provisions for schools that are not included in these categories.*

*Performance Reporting: A Precursor to Determining Rewards and Sanctions*

In most state performance-based accountability systems, performance reporting is a precursor to the determination of rewards and sanctions. O'Reilly (1996, p. 7) explains the theoretical links among performance reporting, rewards and sanctions, and instructional improvement:

> The theory of performance reporting as a means to improvement in student performance results from two assumptions. The first assumption is that schools (or some agent of the schools) will be held accountable for meeting specified performance standards through the application of consequences (either positive or negative) which establishes an incentive for improvement. The second assumption is that information from the performance reporting system will be used to create changes to the teaching and learning process, which will ultimately lead to improvements in student performance. In practice, existing performance reporting systems do not necessarily address these two assumptions adequately.

---

2 This information about Obey-Porter is from the web site of the Northwest Regional Educational Laboratory, found at http://www.nwrel.org/csrdp/about.html.

O'Reilly's explanation clarifies the questions that should be asked about middle-performing schools. First, is continuous improvement of schools accounted for in a State system; that is, can a school receive a reward for improvement relative to its own prior performance, or under value-added, for its own *predicted* level of performance?

States are increasingly offering schools rewards for performance at a certain level, or progress toward a certain level. These rewards may be monetary (like cash awards) or non-monetary (like public recognition). In either case, as Cibulka observes, an important consideration in design is the attractiveness of the "donor's" (in this case, the state's) goal (1989, p. 421). In order for these incentives to work, schools must realize the attractiveness of increasing student performance. Cibulka points to another assumption in performance-based award systems (1989, p. 419):

> It is useful to work within the framework of rational choice theory to understand conditions under which incentives are effective. Incentives are a case of voluntary contractual exchange in which the donor sets forth the terms. Both the donor and the recipient are assumed to be utility-maximizing; they strive to maximize benefits and minimize costs to themselves. Incentives work to the extent that both the donor (in this case, state officials) and the recipients (here, the school, school district, teachers or administrators) perceive that their gains sufficiently exceed their costs so as to justify the voluntary arrangement.

This framework is useful for inference about why a performance-based accountability system may not produce continuous improvements in performance or other outcomes. If the state's "terms" of reward are that schools produce higher student achievement, which will indubitably require setting and maintaining more ambitious curricular goals, schools have to find the cash award sufficient motivation to do so. As Fuhrman wrote in *Rewards and Reform* (1996, p. 332), "what motivates students to learn and teachers to teach involves many strong currents of culture and norms that a program of financial rewards seems a very weak intercession."

Also, it is important to note that the prospect of sanction may not operate as a strong incentive for school improvement. In an accountability system focusing on identifying failing schools (Maryland until recently only ranked schools to determine reconstitution eligibility), many average-performing schools would not experience incentives for improvement. Another state with an accountability system focused on failing schools is New York. Particularly in the past two years, Schools Under Registration Review have become the state's most visible accountability mechanism (other than school report cards), as the Commissioner has placed these schools on probation until they improve.[3] Continuous improvement for other schools is not addressed in a New York statewide accountability system. While the state issues annual report cards (like Maryland), many school leaders say that there is little useful

---

3 Also, low-performing schools, once they have been identified by the state for intervention, may still not experience incentives for improvement. As O'Reilly (1996) writes: "[for low-performing schools] once the threat of sanctions fails to operate as an incentive, there is no theoretical or practical justification to support the notion that state control will lead to improved student performance" (p. 26).

information that would be useful for instructional improvement (Abelmann, Elwell and Lusi, 1997).

Why might it be true that a system that focuses the weight of its incentive bundle on identifying the failures wouldn't improve quality throughout the performance distribution of schools? Some research in the field of public policy and economics illustrates what often occurs when governmental agencies set standards for quality, as in regulation of firms' behavior (Viscusi and Zeckhauser, 1979). Policymakers, the authors explain, often set rigid standards in order to "...promote a higher quality level, assuming the product remains on the market and the firm remains in business" (Viscusi and Zeckhauser, 1979, p. 446). We could substitute "school" for "firm" in that sentence, and in the explanation below, to follow their argument about response to compliance:

> The expense the firm is willing to incur to meet the standard varies predictably with the parameters of the problem. As the probability of inspection or being fined increases, the expected penalty is raised so that it is more likely that the cost of compliance will be lower than the expected cost of noncompliance (Viscusi and Zeckhauser, 1979, p. 446).

Schools in Baltimore with persistently low past achievement will likely find the prospect of reconstitution a serious incentive, and may seek to raise MSPAP assessment scores. But for schools already comfortably above the state threshold for reconstitution, "the shape of the payoff function may be such that the unregulated firm would find it optimal to undertake no quality-enhancing action" (Viscusi and Zeckhauser, 1979, p. 441). Similarly, schools well above the minimum standard will not seek to enhance their quality levels either. These authors' work is also useful because it suggests that governmental action makes assumptions about schools' responses to incentives when increasing quality is the goal; yet schools, like firms, make decisions about whether to improve or not dependent on how they collectively as organizations, perceive "payoffs."

When indicators reveal that a large group of schools are unresponsive, or are failing to show dramatic increases in student scores, it may indicate that school leaders, in Cibulka's formulation, do not perceive that "gains sufficiently exceed their costs so as to justify the voluntary arrangement" (1989, p. 419). Although it is not possible to generalize across all schools, we can think about leaders of schools that are not demonstrating gains. These teachers and administrators may --

1. ...perceive no threat of sanction (i.e., they will not be termed "in decline")
2. ...not be motivated by the potential gain (via the reward), which would require producing change, is worth the costs of producing the change; or they may not believe they can qualify for an award, depending on the program design
3. . ...perceive that the reward is a motivation, but have produced all of the gains they could have, and are lacking the resources or additional motivation to set even higher goals.[4]

---

[4] Clotfelter and Ladd (1996) and Elmore, Abelmann and Fuhrman (1996) have considered these incentive problems. David Cohen has begun to write about the connection between accountability and school capacity.

Carolyn Kelley (1997) conducted research on how the incentives in Kentucky's accountability system had affected schools' performance across the state from cycle 1 to cycle 2, paying particular attention to schools that improved. She found that monetary rewards were not often cited as the most important; desire to avoid negative publicity and the desire to serve students were mentioned far more often.

## II. How Does the Middle Range of Schools Emerge Under Different Accountability Policies? A Look at Mississippi, Kentucky, and Maryland

In a state or district with no accountability policies, statisticians could identify a middle range of schools. Schools within a state or district could be ranked with respect to their socioeconomic status or absolute level of performance on a periodic statewide assessment; we could then say that the middle range was the interquartile range of the 75th and 25th percentiles (with regard to either income level or absolute achievement level).

But under emerging state-level accountability systems, the "middle" is an artifact of state policy, and in order to understand the middle, we must place ourselves within the logic of a given policy's intent and design. Below, I outline three types of emerging systems, offer examples of states within each category, and offer a brief explanation of how a "middle" range can be seen to emerge within each.

### 1. Average test scores or pass rates

*Mississippi: Using Scores to Produce District Rankings.* In Mississippi grades 4 through 9 are tested, and the results are used to produce annual rankings of districts. Districts are ranked along a continuum from 1 (probation) through 5 (excellent). A Level 3 district is one that has met the state's "long-term minimums," which are based on state calculations of a mean performance level. The districts which do not meet the minimum are classified as levels 1 and 2, depending on how far they are from the long-term minimums (CPRE interview with Tom Burnham, Mississippi State Dept., 10/26/94). The majority of districts are ranked at Level 3: on the 1994 state report card, 100 out of the 153 districts were Level 3 (only 4 were "excellent"; the others were "warned" or "probation").

State officials acknowledge that the strongest incentives are directed at the top and bottom: deregulation for levels 4 and 5, and assistance and remediation for levels 1 and 2 (Elmore et al., 1996, p.78). The notable feature of this accountability system is that levels 1 through 3 districts are not competing with each other; they are attempting to get to a defined level of "adequacy" in student performance. However, Levels 4 and 5 are competing against each other for their rankings, since these districts "have to be above the mean of all the districts that are exceeding the long-term minimums to reach level 4 and 5" (Burnham interview, 10/26/94, p.2).

Mississippi's system, then, makes it relatively simple to identify the broad range of schools in the middle range of performance; most of them would be contained in Level 3 districts. Elmore et al. (1996, p. 95) observe that on nationally normed standardized achievement tests, Level 3 districts' scores fall around the thirty-second percentile. The case of Mississippi, then, illustrates how different the "middle" will look from state to state, an artifact of the operative policies. In other states (Kentucky, for instance) schools

in the thirty-second percentile nationally would be targets of state-level incentives to improve. Continuous improvement of teaching and learning is not addressed in Mississippi; the policy goal is to exert pressure for improvement on the poorest-performing districts and help all schools to achieve the adequate level that Level 3 districts represent (i.e., Clune, 1994). It is clear, however, that the Level Three outcomes are not particularly high.

Therefore, the case of Mississippi's accountability system provides an example of how local context underlies policy priorities. Mississippi's schools have a record of low student achievement and low investment in education. Therefore, getting the majority of schools above the failing range is the state's acknowledged priority. One policymaker in Mississippi said of Level 3 schools (in Elmore et al., 1996, p.80): "[They] are achieving a minimum and it is up to the local community to force them to go above the minimum. If the local community is satisfied with the minimum, the state is satisfied with it. The state has limited resources and limited staff, so they have to concentrate on those who are below 3." Under this model, there are virtually no incentives for continuous improvement except for the lowest-performing districts.

*Maryland: Measuring School Effectiveness by State-Established Standards.* The Maryland State Board of Education has used the Maryland Statewide Performance Assessment Program (MSPAP) to set school-level achievement goals (elementary and middle schools only). Since the assessments are graded in each subject area for each student on a scale of 1 to 5, measuring proficiency, the Board has specified how schools are classified based on the percentage of students attaining 3, or proficiency. An important designation the state has made is that a school has reached the "Satisfactory" level when 70% or more of its students have scored 3 or higher in all six content areas of the assessment battery. These assessment scores, along with school attendance information, are used to calculate the School Performance Index, or SPI. Maryland deliberately does not take socioeconomic variables into account when calculating the SPI. Therefore, while Maryland also looks at a Change Index or CI (see below), the SPI can be looked at as a classic use of average test scores or pass rates.

Looking just at the School Performance Index, we could begin to conceptualize a middle range of performance in two different ways. If we were to trust the state's definition of "Satisfactory" for school performance on the MSPAP, we might think of the middle as every single school in the state that is not meeting the state's definition of "reconstitution eligible." If the "satisfactory" designation is a measure of adequate performance (as measured by these particular outcomes on the assessments), then most schools in Maryland fall into a broad middle range. In 1998, there were twenty-three schools where 70% of students' performance was "satisfactory" in grade 3; 8 schools in grade 5; and no schools meeting that standard in grade 8. According to the state, however, there are many schools that are ten points or fewer away from receiving that designation (C. Rosenberger, Maryland State Department of Education, 4/99).[5] Will the state's financial incentives for improvement be enough to induce those schools to improve?

---

5 Eight Maryland schools were designated "excellent," however, based on MSPAP performance: two at grade three and 6 at grade 5 (Maryland State Department of Education).

A second way to derive a middle range of schools from the SPI would be to rank them from highest to lowest, and then choose some statistical measure of central tendency – for instance, specifying that the middle was two standard deviations from the mean, or all schools in the 25-75% interquartile range of SPI scores.

**2. Rates of Improvement Relative to Target Rates of Improvement (or Improving Relative to a Prior Performance Level)**

*Kentucky - Improvement Relative to Target Rate of Improvement* - Under the Kentucky Instructional Results Information System (KIRIS), each school is held accountable based on its progress from a baseline value (established in 1991-92) and a standard representing 100% of students performing at the "proficient" level. The state uses several types of assessments, including performance-based assessment and portfolios, for deriving a multiple index (Elmore, Abelmann and Fuhrman, 1996). Every year, schools are evaluated annually against its threshold and are categorized: reward, successful, improving, in decline, or in crisis. Schools that show improvement over a two-year period are eligible for financial rewards (O'Reilly, 1996).

It has been noted that in this sort of policy design, continuous improvement is accounted for. "In Kentucky, although the system is designed to encourage growth in schools at all performance levels, only exceptional growth is rewarded and only exceptional lack of progress is penalized" (Elmore, Abelmann and Fuhrman, 1996, p. 80). *Under this continuous improvement model, progress in the middle is accounted for; but the other problem of weak incentives for continuous improvement presents itself.*

One way to define the middle would be to look across all schools in the state and identify the schools whose rates of improvement were clustered around the state average rate of improvement (derived from over the nearly eight years of reform). And under another conception, we might look at the middle in Kentucky as those schools who were demonstrated neither notable decline nor growth from cycle one to cycle two (or in future cycles). After all, as the authors cited above note, many members of the education community and the public question what the educational wisdom of having schools continuously improve every biennium over a twenty-year period (Elmore et al., 1996, p. 76). Or the next cycles of the reform may show that schools which showed growth during the first two cycles may have "maxed out" and made all the gains they could with the instructional capacity and leadership they had. Finally, we could still rank Kentucky schools in absolute terms, looking at KIRIS assessment results and pointing out a middle range of achievement; it could then be interesting to look at how the schools had fared under the reward and sanction scheme.

*Maryland's Change Index (CI) - Rank-Ordering a Reconstitution List* - The Change Index is calculated based on current and past two years' School Performance Indices. Schools that have declined are on a list for reconstitution eligibility; schools that have improved for two consecutive years become eligible for cash reward. Since the state does not have the capacity to reconstitute every school on this list, it looks at other contextual factors.

After nearly a decade of this accountability program, not a single school has been reconstituted. Rather, State Superintendent Grasmick selectively engages in negotiation with failing schools. Generally, she and the state board have approached a district with

one or more such schools, and demanded the district's plan for boosting school performance. Often these "negotiations" have resulted in districts firing principals or making other changes in school leadership.

**3. Effectiveness as Measured by Value-Added**

As I noted above, the middle is accounted for in the theory of value-added policies, for instance, the policy instituted several years ago in North Carolina. Policymakers are attempting to measure "how well schools contribute to the learning of their students," as measured against specified curricular outcomes (Ladd et al., 1997). Since the state is setting targets of performance for individual schools, there is no uniform ceiling toward which all schools are progressing (as there is in Maryland, for instance, with progress measured by the SPI).

Yet the state will still face the problem of setting performance targets for individual schools when it seeks to develop a system. The state may not be able to justify the educational logic of the plan. Policymakers under a value-added system might adopt the Kentucky system of setting continuous improvement goals for individual schools over, say, a 20-year period. But incentive problems for schools at high-performing, average-performing, and low-performing are still present. As Ladd et al. note (1997, p. 5):

> It is likely that the targets will be easier to meet in some schools than in others. For schools that start near the 20-year goal, the schools will be deemed effective – and the principals and teachers rewarded – simply for continuing what they were doing in the past whether or not the school's value added is high. Schools starting with low-performing students may find it difficult, if not impossible, to meet their above-average growth targets unless the school is provided with sufficient additional resources and technical assistance to make the target feasible.

So a broad middle could still emerge under a value-added accountability system. Schools with average performance may still have limited capacity to continuously improve toward their goal (based on predicted performance). Still, the benefit of a value-added system is that if implemented with accurate data about inputs and school context, policymakers could utilize valuable information about inputs and capacity. The link between accountability and capacity would be at the forefront of the system's design.

## III. Looking at Performance-Based Accountability Differently

To think about whether and how the middle range of schools should be required to improve is important, because it raises the larger question about what the state role is generally in spurring school improvement. In states with scarce resources for technical assistance, like Mississippi, or a tradition of strong local control, like Vermont, state personnel and citizens alike may not believe that the state can exert much control in matters of continuous improvement. On the other hand, it is clear from my opening analysis that the state policies included this assumption because equity of outcomes is considered an important public good. From this perspective, states must be prepared to continue to develop and improve *all* schools across a range of performance.

In this closing section, I assume the latter: that states, having set these systems in motion, should consider supporting all schools' continuous improvement. I suggest three possible strategies for doing so. These are differentiating and targeting policies; developing and implementing state policies to build capacity (which includes gathering better contextual information about how the policies are affecting different kinds of schools); and investing in the development of shared professional visions of instruction in a standards-based system.

I outline each of these three briefly below.

**1. Differentiating and Targeting Policies**

A clear case study of how an instructional leader recognized the differences in school performance, and then targeted resources differently across the performance distribution, is described by Elmore and Burney (1997) in their description of superintendent Alvarado in District Two. Alvarado, they found, relied on data about school-level performance to make decisions about which schools to focus on most intensively. Rather than holding schools to different standards, he stayed the course in getting all schools to progress toward common performance targets, but his overall strategy recognized their initial differences in starting places.

*Could Incentives be Better Targeted by Changing Policy Design?* - In Mississippi, as we have seen, the state has limited financial resources to offer incentives for districts classified as 3 to move to a 4 or 5 level (Elmore, Abelmann and Fuhrman, 1996). Yet the state's failure to target any incentives at all to level 3 districts is inefficient. Over time, a system with no incentives for excellence will probably cease to produce examples of it, except in the wealthiest communities.

The examples in the paper lead us to think about *how* state policies might be redesigned or "fine-tuned" to either address or spur continuous improvement. Mississippi, for instance, is considering whether to change its accountability system such that districts would be held accountable for improvement (Miller, 1999). In other words, the legislature and the state department might devise incentives targeted specifically for Level 3 districts' progression. Maryland could redirect incentives so that avoiding the state's reconstitution eligibility list is not the focus for schools; perhaps in this case, the state could encourage its counties to set performance goals for individual schools' progress.

There are other examples of states and districts which attempting to build continuous improvement more into the design. For instance, Texas holds schools accountable for the performance of sub-groups of students. Kentucky requires that 10 percent of students in each school move from the lowest performance category (novice) to the next highest (apprentice) in order for the school to be designated "exemplary" in a given reward cycle (Goertz and Chun, 1998).

*Gathering Better Contextual Information –Learning How the Instructional Policies are Landing*

A first step toward understanding how policies might be targeted at a middle range of schools would be to find out more about how accountability measures are affecting schools, or landing. Under the British inspectorate model of accountability, for

instance, the gathering of contextual information about schools is the government's core function. Governmental inspectors make judgments, but their school visits provide a context for understanding up-close what the school's instructional and leadership strengths and shortcomings are.

> Inspectors seek to construct a snapshot of the school at the time of the inspection. We normally think context relates only to the present. But an inspector is able to consider both the past and future of the school. She has access to the vague currents that reveal what the school has been and what it is becoming....an LEA inspector said, `To make a valid judgment about how good a class is now, you need to know what happened before, what the teacher's intentions were, and what happened afterward. You learn about all of these when you visit.' (In Wilson, Reaching for a Better Standard, 1996, p. 120)

Accountability systems are largely driven by assessment results, in combination with selected other outcome indicators. But researchers (i.e. Bryk and Hermanson, 1990) have argued that better organizational indicators and contextual information would help policymakers to understand how instructional policies are affecting schools. For instance, aggregated building-level achievement scores can not tell policy-makers about how various sub-groups of students are performing or improving within a school. As Goertz and Chun (1999) have pointed out, most state accountability policies, by utilizing mostly school-level indicators, often overlook achievement gaps *within* schools. The consequence for equity is that the need for progress of poor and minority students may be masked.

School-level reformers agree that school and district capacity to support change should be deliberately assessed. Phil Schlecty's Center for Leadership in School Reform has attempted to differentiate between standards for teachers and students, and standards for schools. Standard-bearer schools are charged with implementing systemic standards; meanwhile, participating districts open themselves to "...a series of audits and assessments to determine the extent to which the district currently has in place policies, procedures, programs and practices that make it likely that a major restructuring effort can and will be supported and sustained" (Schlecty and Cole, 1992, p. 49). This kind of examination of schools' and districts' policies and governance – standards for schools and communities, as Schlecty and Cole term it -- may provide information to policymakers beyond annual assessment results.

States can also begin to look carefully at what its districts require of schools. While state policy may envision continuous improvement for schools, it may do so while districts still regulate moderately or heavily. This is a promising area for further policy research.

These authors envision that indicators capturing such contextual detail could provide clues to states about what kinds of policies support continuous improvement. "The purpose of the standard-bearer school is to signify the direction reform is taking in school districts with which CSLR has established partnerships. Unlike the pilot school or the model school, the standard-bearer school does not stand apart from the other schools in the district. Rather, the standard-bearer school should belong to all schools in the district" (Schlecty and Cole, 1992, p. 49).

Similarly, a value-added accountability system makes a step in this direction by beginning to differentiate between elements that students contribute and that schools contribute to educational outcomes. Part of the promise of these systems is that contextual information could be used to understand why instructional policy would spur improvement in some contexts and not others.

**2. State Policies to Build Capacity**

Most often, the middle range of schools and districts are held accountable for improvement planning. For example, Florida and Maryland require all schools that have not met state performance standards to write school improvement plans (Massell, 1998). However, this kind of accountability for processes and planning may be limited in the results it can achieve for all schools. A state like Maryland may find that its twin assumptions of making information available to school leaders and letting communities put pressure on may not be sufficient to encourage continuous improvement.

We have seen that many policies are weighted toward identifying and improving low-performing schools. But as Elmore and Fuhrman observed (above), regulation is not the only instrument available to states, nor is a program of financial rewards: there are also capacity-building policies. Once states identify a non-progressing range, it may encourage officials to consider alternative policies that will support learning and teaching.

The interventions that may be most appropriate for schools that are failing to make continuous progress toward performance goals are those directed to improving the quality of classroom instruction. For instance, Massell (1998, p.6) notes that setting professional development standards, changing licensing requirements, or making efforts to bring teachers into the development of curriculum and assessment are all examples of activities to change teachers' knowledge, skills and dispositions. Investing in technology, reducing class size, and supporting teachers' professional networks are other examples of strategies for building school and district capacity (Massell, 1998). Also, states can seek to change the mechanisms for allocating state and federal funds, which may be even easier with the recent passage of federal Education-Flex provisions.

Most states are limited in their capacity to conduct research about program effectiveness. But as more information becomes available about the effectiveness of research-based evaluations, they can play a role in disseminating this information to all schools, not just low-performing ones. There has been much attention paid to adoption of comprehensive school reform models for high-poverty schools (i.e., Ross, Alberg, and Nunnery, 1999; Slavin, 1999), but research-based information about such programs should be shared with middle and higher performing schools, as well.

**3. Develop Shared Visions of Instruction in Standards-Based System**

*"For most educators, the world of the school is a world of particularities, rather than systemic ideas about practice and performance"* (Elmore and Burney, 1997, p. 11).

The caveat about a test-based accountability system is that the results need to become meaningful and useful to educators. A long-term strategy to support an assessment-based accountability system is the development of a language and

understanding about what good teaching practice is, and what quality of student work meets the standards. Some of the best examples of this professional agreement about practice are international (Stevenson and Lee, 1997). In this country, two excellent examples are the National Council of Teachers of Mathematics standards, which have been widely adopted by practitioners; and on a smaller scale, the National Board for Professional Teaching Standards' certification for teachers. These share the outlining of visions of common practice in a standards-based system.

Many of the authors cited in this piece are trying to tackle this problem, though from a variety of perspectives. For instance, Schlecty created a set of principles for "Standard-Bearer Schools," and his Center for Instructional Leadership disseminates information about the top schools' practices. Wilson, in *Reaching for a Better Standard*, argues that the British inspectorate model of accountability is so successful precisely because there is agreement about what aspects of professional practice and school quality the inspectors will evaluate. Mohrman and Wohlstter (1994) identified conditions under which site-based management led to improved teaching and learning. And Elmore (1996) has written about the importance of a common language of practice for sustaining and scaling up instructional reforms. Most educational reforms of this century have been neither sustained nor spread because the reformers did not learn how to share knowledge about changing the core of instruction.

If policy-makers can conceptualize accountability more broadly than assessment results, then the supports and investments to improve instructional practice would follow. Among other likely benefits would be enhancing the credibility of external policies with educators.

**Works Cited:**


Abelmann, Elmore, Even, Kenyon and Marshall. When Accountability Knocks, Will Anybody Answer? University of Pennsylvania: Consortium for Policy Research in Education, 1999.

Abelmann, Elwell, and Lusi. "Plane Crashes, Failing Schools and the Dangers of Not Looking Inside the `Black Box.'" Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, March, 1997.

Bryk, A.S. and Hermanson, K.L. "Educational Indicator Systems: Observations on Their Structure, Interpretation, and Use." Review of Research in Education (vol.19). Washington, DC: American Educational Research Association, 1990.

Burnham, Tom. Mississippi State Commissioner of Education. Telephone interview for the Consortium for Policy Research in Education, 10/26/94.

Cibulka, James. "State Performance Incentives for Restructuring: Can They Work?" Education and Urban Society, Vol. 21, No. 4, August 1989, pp. 417-435.

Clune, William. "The Shift from Equity to Adequacy in School Finance." Educational Policy, Vol. 8, No. 4, pp. 376-94, December 1994.

Cohen, D.K. "Rewarding Teachers for Student Performance." in Rewards and Reform, ed. Susan Fuhrman and J. O'Day. San Francisco, CA: Jossey-Bass, 1996, pp. 60-112.

Committee for Economic Development, Putting Learning First. New York: CED, 1994.

Elmore, R. "Getting to Scale With Good Educational Practice." Harvard Educational Review, Vol 66, No. 1, pp. 1-26, Spring 1996.

Elmore, R., and Burney, D. "School Variation and Instructional Improvement in Community School District #2, New York City." University of Pittsburgh Learning Research Development Center: High-Performance Learning Communities Project, 1997.

Elmore, Abelmann and Fuhrman. "The New Accountability in State Education Reform: From Process to Performance," in Holding Schools Accountable, ed. Helen Ladd. Washington, DC: Brookings, 1996.

Fuhrman, S. and R. Elmore, "Ruling Out the Rules: The Evolution of Deregulation in State Education Policy." Rutgers University, Consortium for Policy Research in Education, 1992.

Fuhrman, Susan. "Conclusion: Building a Better System of Incentives." Rewards and Reform, ed. Fuhrman and O'Day. San Francisco, CA: Jossey-Bass, 1996, pp. 330-341.

Goertz, Margaret, and Chun, T. *Title I and State Education Policy: High Standards for All Students?* Hard Work for Good Schools: Facts Not Fads in Title I Reform, ed. G. Orfield and E.H. DeBray. Harvard University: The Civil Rights Project, 1999.

Goertz, Floden and O'Day, Studies of Education Reform: Systemic Reform, Volume I: Findings and Conclusions. Rutgers University: Consortium for Policy Research in Education, July 1995.

Kelley, Carolyn. "The Kentucky School-Based Performance Award Program: School-Level Effects." Paper presented at AERA in Chicago, IL, March 1997.

Ladd, Helen F., Becky Roselius and Randall Walsh. "Using Student Test Scores to Measure the Effectiveness of Schools." Paper presented at APPAM conference, Washington, DC, 1997.

Making Standards Matter 1997. American Federation of Teachers, 1997.

Massell, D. State Policies for Building Capacity in Education: Progress and Continuing Challenges. University of Pennsylvania: Consortium for Policy Research in Education, 1998.

McDonnell, Lorraine M. and Elmore, R.F. "Getting the Job Done: Alternative Policy Instruments." Educational Evaluation and Policy Analysis, Summer 1987, Vol. 9, No. 2, pp. 133-152.

Miller, Julie. Keeping Their Promise: A Fragile Deal. Mississippi profile in Education Week, *Quality Counts: Rewarding Results, Punishing Failure*, January 11, 1999.

Mohrman and Wohlstetter, "Introduction: Improving School Performance." In School-Based Management: Organizing for High Performance, ed. Mohrman, Wohlstetter and associates. San Francisco: Jossey-Bass, 1994.

National Governors Association, Time for Results. Washington, DC: NGA, 1991.

Newmann, Fred and Bruce King, and Mark Rigdon, "Accountability and School Performance: Implications from Restructuring Schools." Harvard Educational Review, 1997.

O'Reilly, Fran E. "Educational Accountability: Current Practices and Theories in Use." Paper prepared for the Consortium for Policy Research in Education, 1996.

Ross, S. Alberg, M., and Nunnery, J. *Selection and Evaluation of Locally Developed Versus Externally Developed Schoolwide Programs.* Hard Work for Good Schools: Facts Not Fads in Title I Reform, ed. G. Orfield and E.H. DeBray. Harvard University: The Civil Rights Project, 1999.

Rowan, Brian. "Standards as Incentives for Instructional Reform," in Rewards and Reform, ed. Susan Fuhrman and Jennifer O'Day. San Francisco, CA: Jossey-Bass, 1996.

Schlecty, Philip, and Robert Cole. "Creating Standard-Bearer Schools." Educational Leadership, November 1992, pp. 45-49.

Slavin, R.J. *How Title I Can Become the Engine of Reform in America's Schools.* Hard Work for Good Schools: Facts Not Fads in Title I Reform, ed. G. Orfield and E.H. DeBray. Harvard University: The Civil Rights Project, 1999

Steinke, Richard. Director of School Improvement Services, Maryland State Department of Education. Phone conversation of November 4, 1997.

Stevenson, H. and Lee, S. International Comparisons of Entrance and Exit Examinations: Japan, United Kingdom, France, and Germany. Washington, DC: United States Department of Education, 1997.

Summers and Johnson, " The Effects of School-Based Management Plans," in E. Hanushek, editor, Improving America's Schools: The Role of Incentives. Washington, DC: National Academy Press, 1996.

Viscusi and Zeckhauser. "Optimal Standards with Incomplete Enforcement." Public Policy, Vol. 27, No. 4, Fall 1979.

Wilson, Thomas. Reaching for a Better Standard. New York: Teachers College Press, 1996.

Wohlstetter, Priscilla, Smyer, R. and Mohrman, S. "New Boundaries for School-Based Management: The High-Involvement Model." Paper presented at the annual meeting of the American Educational Research Association, April 1994, New Orleans, LA.